Cambio en el grado de inclusión en un modelo multidimensional

Change of degree of containment in a multidimensional model

Francisco Moreno[1], Iván Amón[2], Fernando Arango[1]*

[1]Escuela de Sistemas, Universidad Nacional, Carrera 80 N°. 65-223, Medellín, Colombia

[2]Grupo de investigación GIDATI, Universidad Pontificia Bolivariana, Circular 1 N°. 70-01 Medellín, Colombia

(Recibido 9 de marzo de 2009. Aceptado el 15 de febrero de 2010)

Resumen

Las bodegas de datos usualmente se modelan de forma multidimensional. Los modelos multidimensionales poseen dimensiones las cuales se componen de niveles organizados jerárquicamente de acuerdo con su *inclusión total*. Por ejemplo, en una dimensión geográfica, con niveles Departamento y País, un departamento está incluido totalmente en un país. Recientemente, se ha propuesto una generalización de la inclusión total, la *inclusión parcial*. Por ejemplo, una autopista puede estar incluida sólo en un 20% en un departamento. Sin embargo, ninguno de los trabajos examinados soporta el cambio en el porcentaje de inclusión a través del tiempo. El aporte principal de este artículo es extender un modelo multidimensional con inclusión parcial para soportar este tipo de cambio. La extensión también se incorpora en un lenguaje de consulta multidimensional, lo que permite la formulación de consultas hipotéticas del tipo ¿qué pasaría si?, ¿qué hubiera pasado si?, que pueden ayudar en la toma de decisiones. Para ilustrar la conveniencia de la propuesta se presenta un ejemplo relacionado con accidentes de automóviles.

----- *Palabras clave*: Modelos multidimensionales, bodegas de datos, inclusión total, inclusión parcial, temporalidad

Abstract

Data warehouses are usually modelled in a multidimensional way. The multidimensional models have dimensions composed by hierarchically organized levels according to their *full containment*. For example, in a geographical dimension with Department and Country levels, a department is fully contained into one country. Recently, a generalization of full containment

* Autor de correspondencia: teléfono: + 57 + 4 + 415 90 95, fax + 57 + 4 + 411 23 72, correo electrónico: ivan.amon@upb.edu.co (I. Amón)

has been proposed. It is known as the partial containment. For example, only a 20% of a highway could be contained into a department. In this paper we adopt a multidimensional model that supports partial containment. Our main contribution is to extend this model in order to support the change of the percentage of containment, because the percentage can change over time. To the best of our knowledge, this topic has not been examined in previous works. Our extension is also incorporated into a multidimensional query language, which enables what-if analysis in order to help decision-makers. In order to illustrate the improvements of our proposal, we present a study case related to car accidents.

----- *Keywords:* Multidimensional models, data warehouses, full containment, partial containment, temporality

Introducción

En los últimos años diversos autores [1-8]; han propuesto modelos multidimensionales, usualmente empleados para el modelamiento de bodegas de datos (*data warehouses*) [9, 10]. Estos modelos comparten un conjunto de conceptos esenciales como dimensión, jerarquía, nivel, hecho, medida, entre otros. Un modelo multidimensional posee dimensiones, por ejemplo la dimensión Tiempo y la dimensión Geográfica, que se asocian con un fenómeno medible de interés para una organización, denominado hecho, por ejemplo accidentes de automóviles. Una dimensión representa una perspectiva del negocio para analizar los hechos y se compone de un conjunto no vacío de niveles (Día, Mes y Año son niveles de la dimensión Tiempo; Autopista, Departamento y País son niveles de la dimensión Geográfica).

Los hechos poseen medidas, indicadores del comportamiento de una determinada actividad de la organización [11], por ejemplo número de accidentes, número de víctimas; sobre las cuales se enfocan los cálculos y los informes en una organización. Los niveles de una dimensión se organizan jerárquicamente de acuerdo con las necesidades de análisis de la información [12]. La relación jerárquica entre los niveles captura su *inclusión total*. Por ejemplo, en la dimensión Geográfica, un departamento está incluido totalmente en un país. Recientemente Jensen [7] propuso una generalización de la inclusión total, la *inclusión parcial*.

La inclusión parcial permite representar situaciones en las que un valor de un nivel no está incluido totalmente en otro. Por ejemplo, una autopista puede estar incluida sólo en un 20% en un departamento. Sin embargo, el modelo de Jensen [7] no soporta el posible cambio en el porcentaje de inclusión. Por ejemplo, en un tiempo t_i el porcentaje de inclusión de una autopista en un departamento es del 20%, pero en un tiempo t_{i+1} este porcentaje puede cambiar debido a la construcción o destrucción de tramos de la autopista. Para dar soporte a este tipo de cambio, en este artículo se extiende el modelo de Jensen. La extensión también se incorpora en un lenguaje de consulta multidimensional.

Este artículo está organizado así: inicialmente se presenta un ejemplo motivador a modo de caso de estudio; luego se presentan los conceptos esenciales de un modelo multidimensional con inclusión parcial. A continuación, se presenta la extensión que sirve de soporte al cambio en el porcentaje de inclusión. Finalmente, se presentan las conclusiones y los trabajos futuros.

Ejemplo motivador

Considérese la infraestructura vial de un país compuesta de autopistas que atraviesan sus departamentos. La figura 1 ilustra una situación en la cual tres autopistas (Ap_1, Ap_2 y Ap_3), atraviesan tres departamentos ($Dpto_1$, $Dpto_2$ y $Dpto_3$).

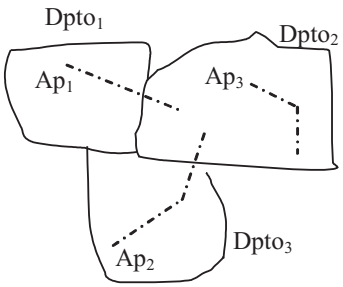


Figura 1 Infraestructura vial de un país

Para las autoridades de tránsito es de interés analizar aspectos como la accidentalidad. Les interesa saber, por ejemplo, cuáles autopistas presentan mayor accidentalidad para mejorar su control, modificar su trazado o tomar otras medidas con el fin de disminuir la accidentalidad. En este escenario, los accidentes son los fenómenos de interés (es decir son los hechos) los cuales ocurren en un lugar y en una fecha determinados (dimensiones geográfica y temporal). En la figura 2 se presenta el modelo multidimensional correspondiente (se usa la notación de Jensen [7]) y en la tabla 1 se presenta una muestra de la tabla de hechos de accidentes. Nótese que cada hecho corresponde al conjunto de accidentes que ocurrieron en una autopista en una fecha en particular.

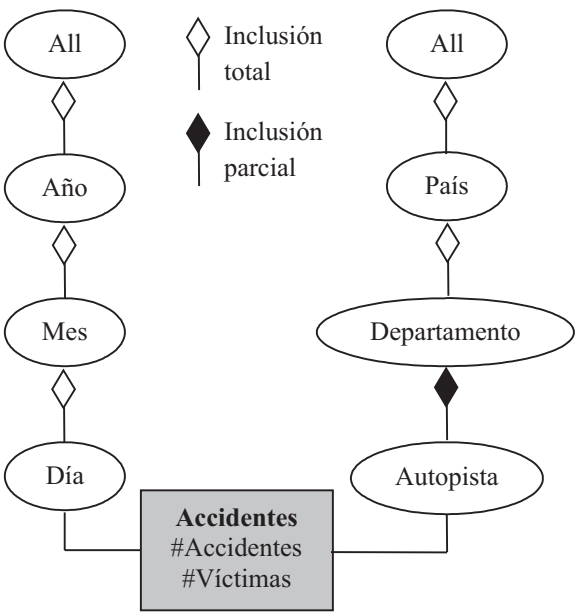


Figura 2 Modelo multidimensional para el análisis de accidentes

Tabla 1 Muestra de datos de la tabla de hechos de accidentes

Dimensiones		*Medidas*	
Día	*Autopista*	*#Accidentes*	*#Víctimas*
	...		
01/01/2008	Ap_1	2	5
01/01/2008	Ap_2	1	2
02/01/2008	Ap_1	3	9
02/01/2008	Ap_2	1	2
03/01/2008	Ap_3	1	3
04/01/2008	Ap_2	2	4
	...		
20/01/2008	Ap_2	3	3
	...		

Supóngase que el porcentaje de inclusión de la autopista Ap_2 en el departamento $Dpto_2$ es del 20% y en el departamento $Dpto_3$ es del 80%. Sea la consulta: ¿Cuál es el número total de accidentes que han ocurrido en el departamento $Dpto_2$?

A partir de la figura 1, se observa que los hechos asociados con la autopista Ap_3 contribuyen con el total solicitado ya que dicha autopista está incluida totalmente en el departamento $Dpto_2$; sin embargo con respecto a los hechos ocurridos en la autopista Ap_2 no se tiene tal certeza.

No obstante, es posible dar una respuesta aproximada a esta consulta si se considera el porcentaje de inclusión de una autopista en un departamento y se distribuyen proporcionalmente los datos como se muestra en la tabla 2.

Supóngase ahora que el porcentaje de inclusión de la autopista Ap_2 en los departamentos $Dpto_2$ y $Dpto_3$ cambia como se muestra en la figura 3. El porcentaje de inclusión de la autopista Ap_2 en ambos departamentos es ahora del 50% debido a la adición de un tramo a la autopista.

Tabla 2 Cálculo del total de accidentes en el $Dpto_2$ (se considera un porcentaje de inclusión del 20% de la autopista Ap_2 en $Dpto_2$)

Autopista	*Total de accidentes en la autopista*	*% de inclusión en $Dpto_2$*	*Número estimado de accidentes en $Dpto_2$*
Ap_1	5	20	5 * 0,2 = 1
Ap_2	7	20	7 * 0,2 = 1,4
Ap_3	1	100	1 * 1 = 1
	Total		3,4

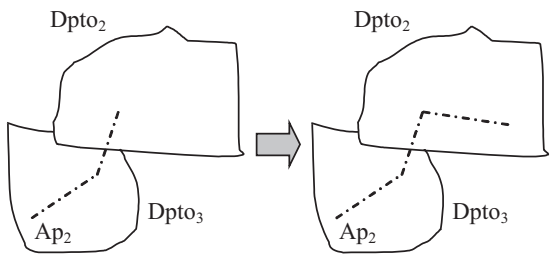


Figura 3 Cambio en la inclusión parcial: crecimiento de la autopista Ap_2

Considérese de nuevo la consulta planteada y supóngase que el nuevo tramo se habilita para el tránsito vehicular a partir del 15/01/2008. Nótese que se debe conservar la evolución de los cambios de los porcentajes de inclusión de las autopistas en los departamentos, con el fin de obtener resultados consistentes con el tiempo. De lo contrario, todos los hechos anteriores al 15/01/2008 asociados con la autopista Ap_2, darían la impresión de que sucedieron cuando el porcentaje de inclusión de la autopista Ap_2 en ambos departamentos es del 50%. La Tabla 3 muestra los resultados a los que conduciría la aplicación del porcentaje de inclusión a todos los datos sin considerar la fecha correspondiente.

De otro lado, los resultados de la tabla 4 son consistentes respecto al porcentaje de inclusión existente en el momento en que sucedieron los hechos.

Tabla 3 Cálculo del total de accidentes en el Dpto2 (se considera un porcentaje de inclusión del 50% de la autopista Ap2 en Dpto2

Autopista	*Total de accidentes en la autopista*	*% de inclusión en $Dpto_2$*	*Número estimado de accidentes en $Dpto_2$*
Ap_1	5	20	5 * 0,2 = 1
Ap_2	7	50	7 * 0,5 = 3,5
Ap_3	1	100	1 * 1 = 1
	Total		5,5

Tabla 4 Cálculo del total de accidentes en el $Dpto_2$ (se considera el porcentaje de inclusión correspondiente a la fecha de los hechos)

Autopista	*Total de accidentes en la autopista*	*% de inclusión en $Dpto_2$*	*Número estimado de accidentes en $Dpto_2$*
Ap_1	5	20	5 * 0,2 = 1
Ap_2	4	20	4 * 0,2 = 0,8
Ap_2	3	50	3 * 0,5 = 1,5
Ap_3	1	100	1 * 1 = 1
	Total		4,3

En el modelo de Jensen [7] no se conserva la historia de este tipo de cambios. En la Sección *Soporte al grado de inclusión*, se presenta la propuesta de extensión correspondiente para soportar esta situación.

Modelo multidimensional

A continuación se presentan los conceptos esenciales del modelo multidimensional de Jensen [7] el cual soporta inclusión parcial.

Esquema multidimensional

Un *esquema multidimensional* es una dupla E = (F, TD), donde F es un *tipo de hecho* y TD = {td_i,

i = 1,…, n} es un conjunto de *tipos de dimensión*. Un tipo de dimensión td es una cuadrupleta (TN_{td}, @, All, ¯), donde TN_{td} = {tn_i, i = 1,…, k} es un conjunto de *tipos de niveles*. @ es un orden parcial en el conjunto TN_{td}. All es el elemento superior (*top*) del orden parcial y ¯ representa el elemento inferior (*bottom*) del orden parcial. All representa el nivel de agrupación más alto de los valores dimensionales y ¯ el más bajo. El dominio de All es un único valor: dom(All) = {*all*}.

Ejemplo 1 Sea el esquema multidimensional Accidentes = {A, TD}, donde A es un tipo de hechos de accidentes y TD = {Tiempo, Ubicación}:

Tiempo = (TN_{Tiempo}, @, All, ¯), TN_{Tiempo} = {Día, Mes, Año, All} y ¯ = Día. El orden parcial correspondiente se muestra en la Figura 4 (a).

Ubicación = ($TN_{Ubicación}$, @, All, ¯), $TN_{Ubicación}$ = {Autopista, Departamento, País, All} y ¯ = Autopista. El orden parcial correspondiente se muestra en la Figura 4 (b).

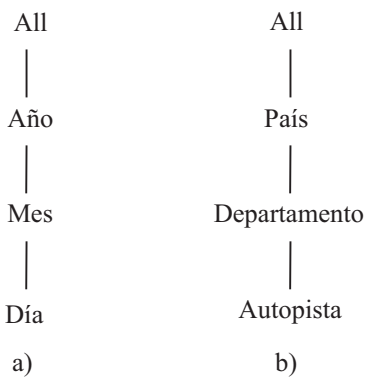


Figura 4 Orden parcial: a) dimensión Tiempo y b) dimensión Ubicación

Nótese que para representar un orden parcial @, se usa su reducción transitiva (Diagrama de Hasse [13]).

Instancia de dimensión

Dado un esquema multidimensional (F, TD), una *instancia de dimensión* d, de tipo td Î TD, es una dupla d = (N_d, §), donde N_d = {n_i, i = 1,…, k} es un conjunto de niveles. Cada nivel n es de tipo tn Î TN_{td}, es decir, un nivel n es un conjunto de valores de tipo tn. § es un orden parcial en È$_j$ n_j (unión de todos los valores de los niveles de una instancia de dimensión). De acá en adelante se escribirá Dim en lugar de È$_j$ n_j.

Ejemplo 2 Sea tiempo una instancia del tipo de dimensión Tiempo y ubicación una instancia del tipo de dimensión Ubicación, véase Ejemplo 1:

tiempo = {N_{tiempo}, §}, N_{tiempo} = {día, mes, año, all}, donde día es de tipo de nivel Día, mes es de tipo de nivel Mes, año es de tipo de nivel Año y all es de tipo de nivel All. día = {01/01/2007, 02/01/2007, …, 31/12/2008}, mes = {01/2007, 02/2007, …, 12/2008}, año = {2007, 2008} y all = {*all*}. El orden parcial correspondiente se muestra en la Figura 5 (a).

ubicación = {$N_{ubicación}$, §}, $N_{ubicación}$ = {autopista, departamento, país, all}, donde autopista es de tipo de nivel Autopista, departamento es de tipo de nivel Departamento, país es de tipo de nivel País y all es de tipo de nivel All. autopista = {Ap_1, Ap_2, Ap_3}, departamento = {$Dpto_1$, $Dpto_2$, $Dpto_3$}, país = {P_1} y all = {*all*}. El orden parcial correspondiente se muestra en la Figura 5 (b).

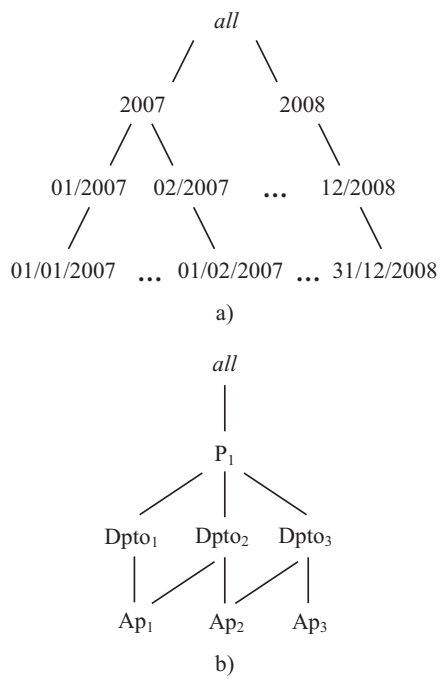


Figura 5 Instancias de dimensión: a) tiempo y b) ubicación

Grado de inclusión

Dados dos valores *a* Î Dim y *b* Î Dim y un número g Î [0; 1], la notación *a* §$_g$ *b* significa que *a* está incluido en *b* en un g*100%. g es el *grado de inclusión* de *a* en *b*. Si g = 1 se dice que *a* está incluido totalmente en *b* y si g = 0 significa que *a podría* estar incluido en *b* (si existe inclusión, se desconoce el valor del grado).

Jensen en [7] presenta algunas reglas de transitividad que permiten inferir grados de inclusión entre valores dimensionales (*c* Î Dim, p Î [0; 1) y q Î [0; 1)):

i) transitividad entre inclusiones totales: si *a* §$_1$ *b* y *b* §$_1$ *c*, entonces *a* §$_1$ *c*,
ii) transitividad entre inclusión parcial y total: si *a* §$_p$ *b* y *b* §$_1$ *c*, entonces *a* §$_p$ *c*,
iii) transitividad entre inclusión total y parcial: si *a* §$_1$ *b* y *b* §$_p$ *c*, entonces *a* §$_0$ *c*,
iv) transitividad entre inclusiones parciales: si *a* §$_p$ *b* y *b* §$_q$ *c*, entonces *a* §$_0$ *c*.

Por ejemplo, la regla iii) establece que si *a* está incluido totalmente en *b* y *b* está incluido en *c* en un p*100% (p < 1), sólo se puede inferir que *a podría* estar incluido en *c* (*a* §$_0$ *c*).

Relación hecho-dimensión

Una relación *Hecho-Dimensión* r se define como r Í f ´ Dim, donde f es un conjunto de hechos de tipo F, véase subsección *Esquema Multidimensional*. Cada hecho debe estar relacionado con al menos un valor de cada dimensión. Por simplicidad, se supondrá que cada hecho se asocia con un único valor de cada dimensión y que el valor dimensional correspondiente pertenece al nivel inferior (*bottom*) de la dimensión.

Ejemplo 3 Considérese de nuevo el Ejemplo 1. Sea accidentes = {Ac_1, Ac_2, Ac_3, Ac_4, Ac_5} un conjunto de hechos de tipo A. Sean las relaciones Hecho-Dimensión:

r_1 = {(Ac_1, 01/01/2008), (Ac_2, 01/01/2008), (Ac_3, 02/01/2008), (Ac_4, 02/01/2008), (Ac_5, 03/01/2008)} y

r_2 = {(Ac_1, Ap_1), (Ac_2, Ap_2), (Ac_3, Ap_1), (Ac_4, Ap_2), (Ac_5, Ap_3)}

Las relaciones r_1 y r_2 asocian el conjunto de hechos accidentes con valores de la instancia de dimensión tiempo y con la instancia de dimensión ubicación del Ejemplo 2, respectivamente.

Caracterización de hechos

El concepto de caracterización de hechos se define a partir de una relación Hecho-Dimensión r. Se dice que un hecho está caracterizado por un valor dimensional, si el hecho se asocia directa o indirectamente (por transitividad en el orden parcial § de los valores de la dimensión) con dicho valor, es decir, un hecho h Î f está caracterizado por un valor v_1 Î Dim, escrito h ® v_1, de una dimensión si: (h, v_1) Î r ó si existe un valor v_2 Î Dim tal que (h, v_2) Î r y v_2 § v_1.

Ejemplo **4** En la figura 6: Ac_1 ® Ap_1, Ac_1 ® $Dpto_2$, Ac_1 ® $Dpto_3$, Ac_1 ® P_1, Ac_5 ® Ap_3, Ac_5 ® $Dpto_2$ y Ac_5 ® P_1.

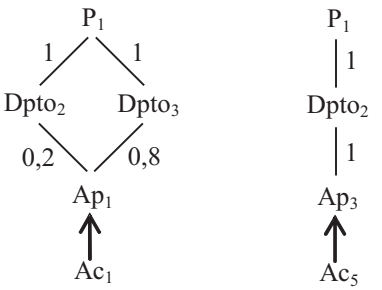


Figura 6 Hechos Ac_1 y Ac_5 asociados con valores dimensionales

Objeto multidimensional

Luego de especificar las dimensiones, la relación Hecho-Dimensión y la caracterización de hechos se define el objeto multidimensional (OM). Informalmente, un OM es un cubo de datos [14], es decir, un arreglo de celdas (que contienen las medidas) asociadas con un conjunto de valores dimensionales. Formalmente, un OM es una cuadrupleta OM = (E, f, D, R), donde E = (F, TD) es un esquema multidimensional, f es un

conjunto de hechos de tipo F, D es un conjunto de instancias de dimensiones cada una de tipo td Î TD y R es un conjunto de relaciones Hecho-Dimensión.

Ejemplo 5. Sea un OM CuboAccidentes = (Accidentes, accidentes, {tiempo, ubicación}, {r_1, r_2}). Donde Accidentes es el esquema multidimensional del Ejemplo 1, accidentes el conjunto de hechos del Ejemplo 3, {tiempo, ubicación} es el conjunto formado por las instancias de dimensiones del Ejemplo 2 y {r_1, r_2} es el conjunto formado por las relaciones Hecho-Dimensión del ejemplo 3.

Soporte al cambio en el grado de inclusión

El grado de inclusión entre dos valores dimensionales puede cambiar a través del tiempo. Por ejemplo, en la figura 3 se representa el cambio en el grado de inclusión entre la autopista Ap_2 y los departamentos $Dpto_2$ y $Dpto_3$.

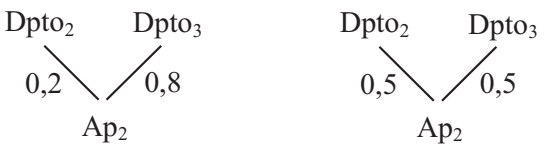
Con el fin de registrar el cambio en el grado de inclusión, se propone la siguiente extensión al modelo de la sección anterior. Sea un tipo de dimensión (TN_{td}, @, All, ¯, m) donde m es una unidad de tiempo (horas, días, meses, años, entre otras). m define la precisión temporal requerida (granularidad) por la aplicación para registrar la evolución del grado de inclusión entre los valores de una dimensión.

Considérese una pareja de tipos de niveles (tn_1, tn_2) Î TN_{td}. Sea una instancia de dimensión d = (N_d, §), de tipo td. Sean los niveles n_1 Î N_d y n_2 Î N_d, n_1 es de tipo de nivel tn_1 y n_2 es de tipo de nivel tn_2. Para la pareja (n_1, n_2) se define una función GI (Grado de Inclusión) con signatura: n_1 ´ n_2 ´ dom(m) à [0;1]. La función GI devuelve el grado de inclusión en un tiempo dado de un valor de n_1 con respecto a un valor de n_2.

Ejemplo 6. Sea el tipo de dimensión Ubicación = ($TN_{Ubicación}$, @, All, ¯, Día). Sea la pareja de tipos de niveles (Autopista, Departamento) del Ejemplo 1. Sea ubicación = {$N_{ubicación}$, §} una instancia del tipo de dimensión Ubicación, $N_{ubicación}$ = {autopista, departamento, país, all}. autopista es de tipo de nivel Autopista y departamento es de tipo de nivel Departamento. Para la pareja (autopista, departamento) se define una función GI; algunos de sus valores se muestran en la tabla 5 y se grafican en la figura 7. Por ejemplo GI(Ap_2, $Dpto_3$, 01/01/2008) = 0,8 y GI(Ap_2, $Dpto_3$, 15/01/2008) = 0,5.

Tabla 5 Muestra de datos de la función GI para (autopista, departamento). ap Î autopista, dp Î departamento y t Î dom(Dí

ap	*dp*	*t*	*GI*
		...	
Ap_2	$Dpto_2$	01/01/2008	0,2
Ap_2	$Dpto_3$	01/01/2008	0,8
Ap_2	$Dpto_2$	02/01/2008	0,2
Ap_2	$Dpto_3$	02/01/2008	0,8
		...	
Ap_2	$Dpto_2$	15/01/2008	0,5
Ap_2	$Dpto_3$	15/01/2008	0,5
		...	



a) b)

Figura 7 Grado de inclusión de la autopista Ap_2 en los departamentos $Dpto_2$ y $Dpto_3$: a) entre 01/01/2008 y 14/01/2008 y b) a partir de 15/01/2008

Para el cálculo del grado de inclusión entre dos valores dimensionales no adyacentes en la jerarquía, se aplican las reglas de transitividad de la subsección *Grado de Inclusión*.

Ejemplo 7 Considérese la figura 1 y supóngase que el GI(Ap_1, $Dpto_1$, 31/01/2008) = 0,8, véase Figura 8(a). Supóngase que a partir del 01/02/2008, el tramo de la autopista Ap_1 en el

departamento $Dpto_2$ se elimina, por lo tanto $GI(Ap_1, Dpto_1, 01/02/2008) = 1$, véase figura 8(b). Por consiguiente, al aplicar las reglas de transitividad, se obtiene que $GI(Ap_1, P_1, 31/01/2008) = 0,8$ y $GI(Ap_1, P_1, 01/02/2008) = 1$.

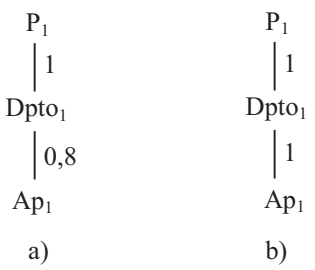


Figura 8 Grado de inclusión de Ap_1 en $Dpto_1$: a) el 31/01/2008 y b) el 01/02/2008

Consultas

En esta sección, se ilustra cómo la extensión propuesta puede ser usada en un lenguaje de consulta multidimensional. Aunque MDX (*Multidimensional Expressions*) [15], es un lenguaje que se ha convertido en los últimos años en un estándar *de facto* para la consulta de datos multidimensionales, en este trabajo se usará el lenguaje propuesto por Datta [16], debido a su similitud con el álgebra relacional. Se usan los operadores de selección (s) y de agrupamiento (a). Para todas las consultas se usa el cubo CuboAccidentes del Ejemplo 5.

Consulta 1. Obtener el número total de accidentes que han ocurrido en el departamento $Dpto_2$.

$$a_{[SUM(\#Accidentes * GI(autopista, `Dpto2', día))]}(\text{CuboAccidentes})$$

Es decir, se seleccionan todos los hechos del cubo CuboAccidentes. Para cada hecho se halla el grado de inclusión de la autopista correspondiente en el departamento $Dpto_2$ y este valor se multiplica por el número de accidentes. Luego mediante la función de agregación SUM se obtiene el total solicitado. La misma consulta expresada en un estilo similar a SQL sería:

```
SELECT SUM(#Accidentes * GI(autopista, 'Dpto2', día))
FROM CuboAccidentes
```

Nótese que para calcular el grado de inclusión se usa la fecha (día) asociada con el hecho. Sin embargo, es posible plantear consultas *hipotéticas* con el fin de analizar comportamientos pasados y realizar pronósticos, como se ejemplifica a continuación.

*Consulta 2***.** ¿Cuál hubiera sido el número total de accidentes en el $Dpto_2$ si se considera la inclusión que existía de las autopistas en dicho departamento el 01/01/2007?

$$a_{[SUM(\#Accidentes * GI(autopista, `Dpto2', `01/01/2007'))]}((\text{CuboAccidentes}))$$

En esta consulta se consideran todos los hechos del cubo CuboAccidentes, por ejemplo hechos del 2007 y del 2008, pero se usa el grado de inclusión correspondiente al 01/01/2007.

Consulta 3. ¿Cuál habría sido el número total de accidentes en el $Dpto_2$ en el 2007 si se considera la inclusión actual de las autopistas en dicho departamento? La fecha actual se representa con un valor *now*.

$$a_{[SUM(\#Accidentes * GI(autopista, `Dpto2', now))]}(s_{día > `01/01/2007' AND dia < `31/12/2007'}(\text{CuboAccidentes}))$$

En esta consulta se seleccionan sólo los hechos del cubo CuboAccidentes del 2007, pero se usa el grado de inclusión correspondiente a la fecha actual.

Conclusiones y trabajo futuro

En este trabajo se adoptó un modelo multidimensional que soporta inclusión parcial y se extendió para permitir el posible cambio en el grado de inclusión entre valores dimensionales.

La extensión también se incorporó en un lenguaje de consulta multidimensional. Esto permite plantear consultas que son consistentes de acuerdo con el tiempo y además permite formular consultas hipotéticas (¿qué pasaría si?, ¿qué hubiera pasado si?), que pueden ayudar a los usuarios en la toma de decisiones.

Como trabajo futuro se planea implementar el modelo propuesto, para ello se podría usar una plataforma como *Pentaho* [17] o *Microsoft Analysis Server* [18], sin embargo, debido a que estas plataformas están orientadas al manejo de modelos multidimensionales que soportan inclusión total, la incorporación de la extensión propuesta plantea desafíos interesantes.

De otro lado, desde el punto de vista del lenguaje, ambas plataformas soportan MDX. Sin embargo, debido a que MDX también se orienta al manejo de la inclusión total, la incorporación de la propuesta en éste, conlleva igualmente dificultades.

Agradecimientos

Este artículo hace parte del Doctorado en Ingeniería - Sistemas de la Universidad Nacional de Colombia Sede Medellín, auspiciado por Colciencias.

Referencias

1. R. Agrawal, A. Gupta, S. Sarawagi. "Modeling Multidimensional Databases". *13th International Conference on Data Engineering (ICDE'97)*. Birmingham. Inglaterra. 1997. pp. 232-243.

2. M. Gyssens, L. Lakshmanan. "A Foundation for Multi-dimensional Databases". *23rd International Conference on Very Large Data Bases (VLDB'97)*. Atenas. Grecia. 1997. pp. 106-115.

3. P. Vassiliadis. "Modeling Multidimensional Databases, Cubes and Cube Operations". *10th International Conference on Scientific and Statistical Database Management* (*SSDBM'98)*. Capri. Italia. 1998. pp. 53-62.

4. M. Golfarelli, S. Rizzi. "A Methodological Framework for Data Warehouse Design". *1st ACM International Workshop on Data Warehousing and OLAP (DOLAP'98)*. Washington D.C. Estados Unidos. 1998. pp. 3-9.

5. W. Lehner, J. Albrecht, H. Wedekind. "Normal Forms for Multidimensional Databases". *10th International Conference on Scientific and Statistical Database Management* (*SSDBM'98)*. Capri. Italia. 1998. pp. 63-72.

6. T. B. Pedersen, C. S. Jensen, C. E. Dyreson. "A Foundation for Capturing and Querying Complex Multidimensional Data". *Information Systems*. Vol. 26. 2001. pp. 383-423.

7. C. S. Jensen, A. Kligys, T. B. Pedersen, I. Timko. "Multidimensional Data Modeling for Location-based Services". *10th ACM International Symposium on Advances in Geographic Information Systems (GIS 2002)*. McLean. USA. 2002. pp. 55-61.

8. I. Timko, C. E. Dyreson, T. B. Pedersen. "Probabilistic Data Modeling and Querying for Location-based Data Warehouses". *17th International Conference on Scientific and Statistical Database Management* (*SSDBM* 2005). Santa Bárbara. USA. 2005. pp. 273-282.

9. W. H. Inmon. *Building the Data Warehouse*. 3ª. ed. Ed. John Wiley & Sons. Nueva York. USA. 2002. pp. 1-432.

10. R. Kimball, M. Ross. *The Data Warehouse Toolkit: the Complete Guide to Dimensional Modeling*. Ed. John Wiley & Sons, Nueva York. USA. 2ª. ed. 2002. pp. 1-464.

11. E. Malinowski, E. Zimányi. *Advanced Data Warehouse Design: From Conventional to Spatial and Temporal Applications*. Ed. Springer, Nueva York. USA. 2008. pp. 1-435.

12. R. Torlone. "Conceptual Multidimensional Models". M. Rafanelli, *Multidimensional Databases: Problems and Solutions*, Ed. Idea Group Publishing, Pennsylvania. USA. 2003. pp. 69-90.

13. R. Freese. "Automated Lattice Drawing". *2nd International Conference on Formal Concept Analysis (ICFCA'04)*. Sydney. Australia. 2004. pp. 112-127.

14. OLAP Council. *The OLAP glossary*. The OLAP Council. 1997. http://www.olapcouncil.org/research/resrchly.htm. Consultada el 6 de Febrero de 2009.

15. M. Whitehorn, R. Zare, M. Pasumansky. *Fast Track to MDX*. 2ª. ed. Ed. Springer, New York. USA. 2006. pp. 1-310.

16. A. Datta, H. Thomas. "The Cube Data Model: a Conceptual Model and Algebra for On-line Analytical Processing in Data Warehouses". *Decision Support Systems.* Vol. 27. 1999. pp. 289-301.

17. Pentaho. *Pentaho BI Suite Enterprise Edition.* http://www.pentaho.com. Consultada el 16 de mayo de 2009.

18. Microsoft. *Microsoft SQL Server 2008*. http://www.microsoft.com/sqlserver/2008/en/us. Consultada el 16 de mayo de 2009.